Beyond Analysis

Team Name: VoidMen

Abishek Mahesh - Prithvi Seshadri - Rishi Dinesh - Pranc

Introduction

Problem Statement

Head digital works has provided a dataset that falls in the domain of online skill-based gaming. This dataset contains a 22-dimensional feature space that describes the differ aspect of customer behaviour. The goal is to predict the Y1 and Y2 values of a given customer, which represents the customer's current value and future value (temporal extrapolation) respectively

Dataset

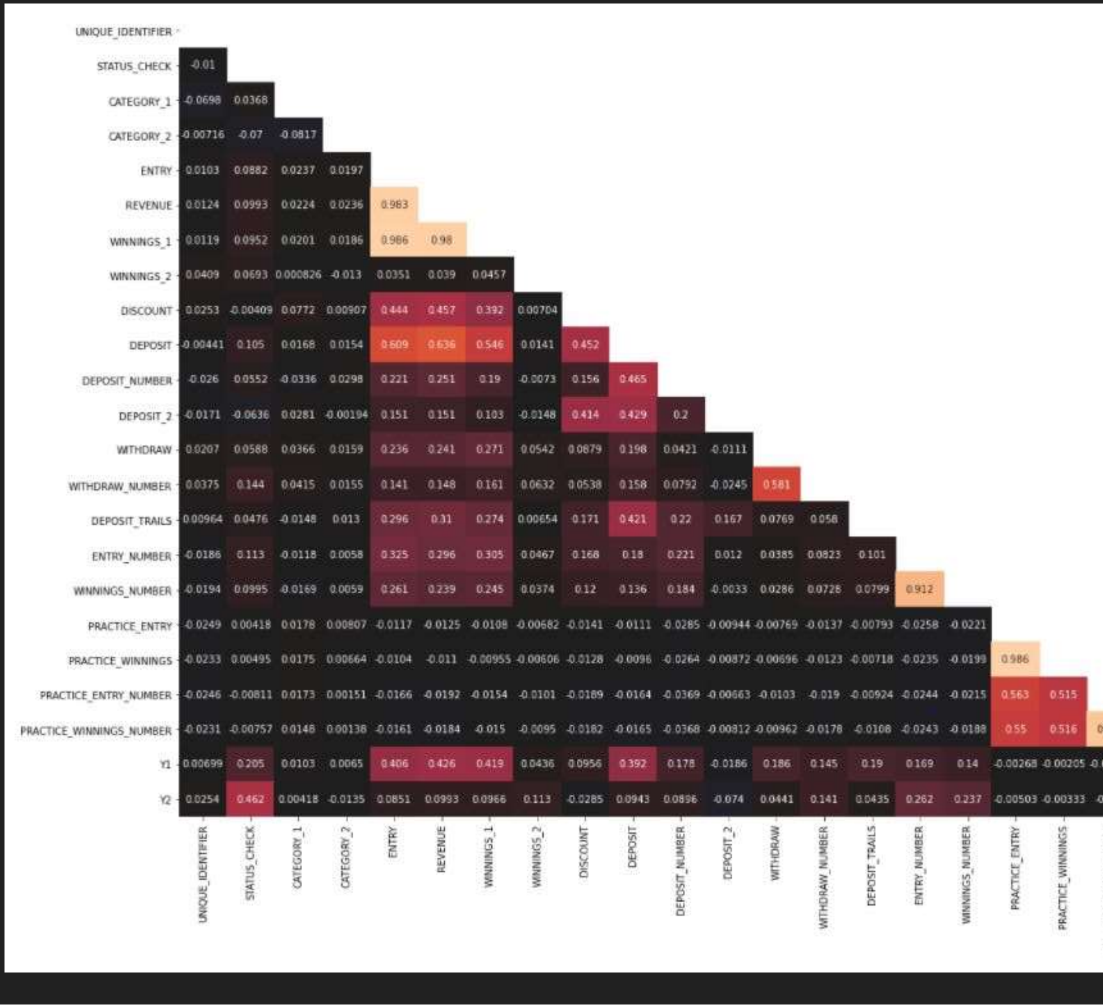
The dataset is divided into train and test sets. Each customer is identified by a UNIQUE_IDENTIFIER and a sequence of entries tagged by the SEQUENCE_NO. Various c information attributing to the performance of the customer in the game has been prov

Metrics

Root Mean Squared Error(RMSE) is the primary metric used in our analysis.

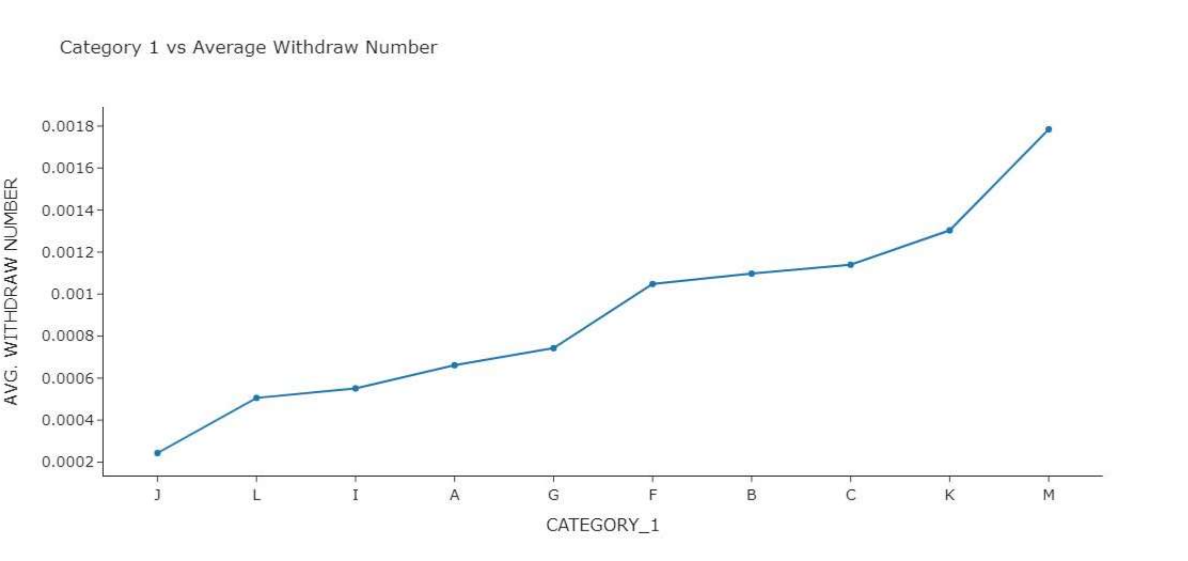
Exploratory Data Analysis

Correlation Heatmap of Numerical Features

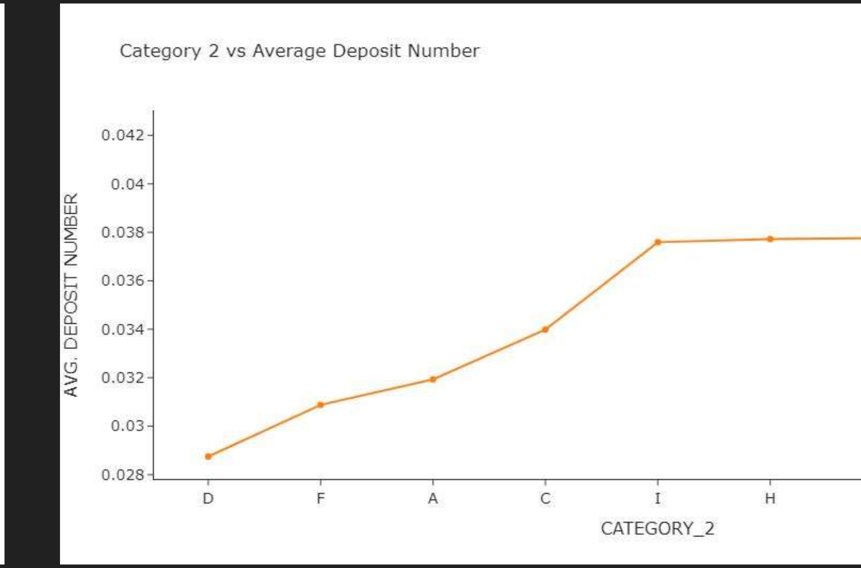


Exploratory Data Analysis

CATEGORY_1 Visualization

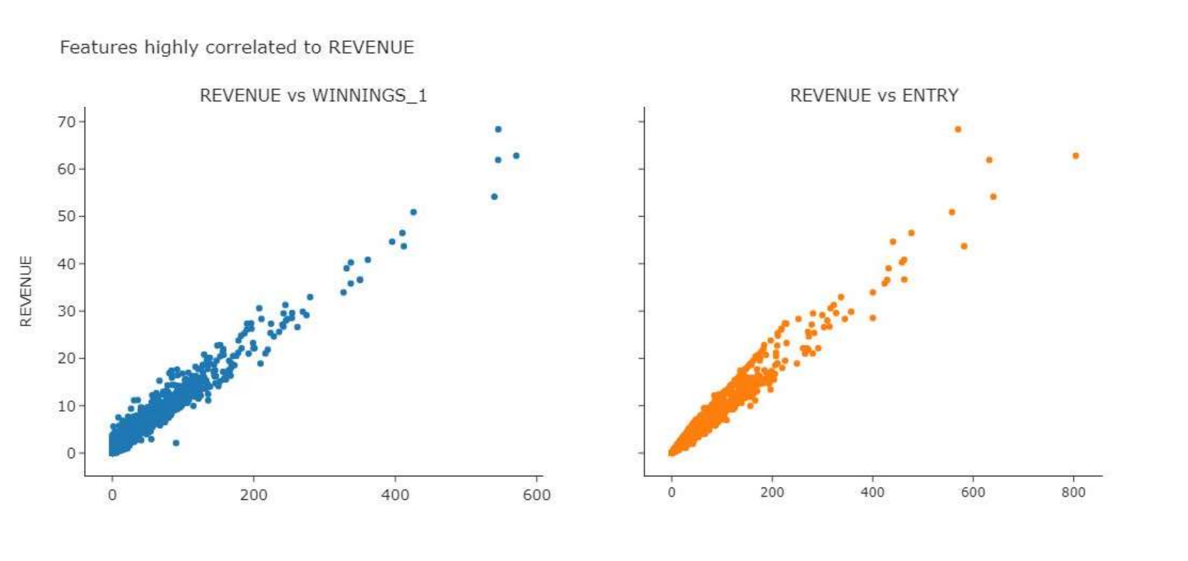


CATEGORY_2 Visual

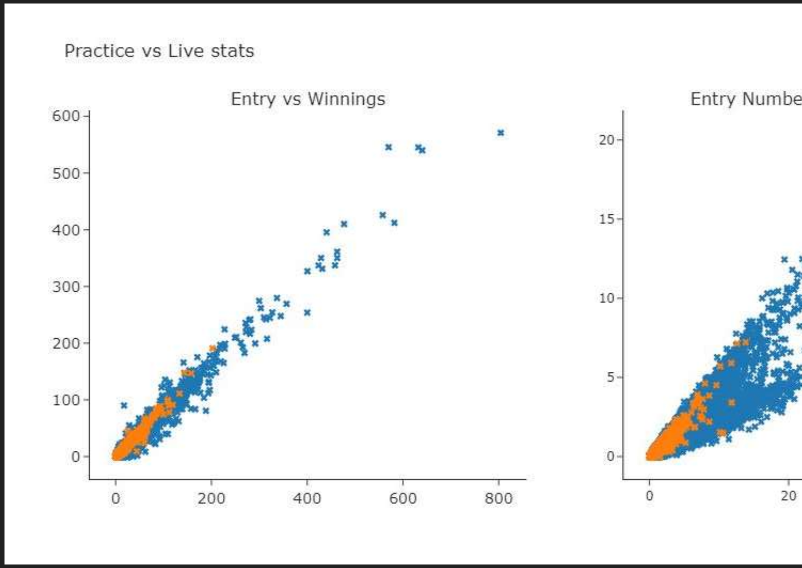


Exploratory Data Analysis

REVENUE Correlation Analysis

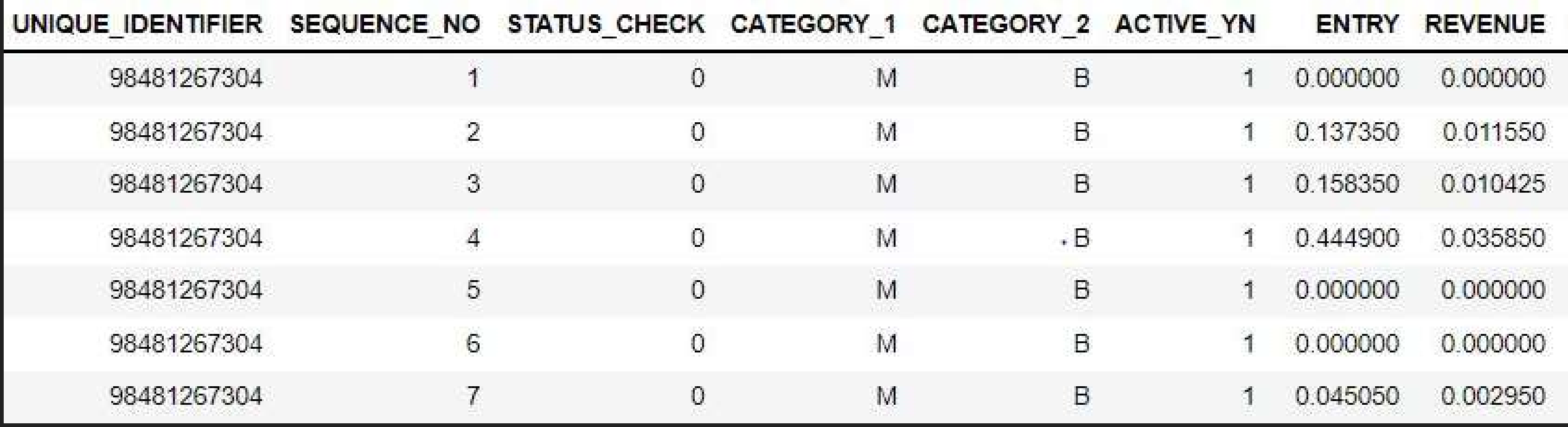


Practice vs Live D

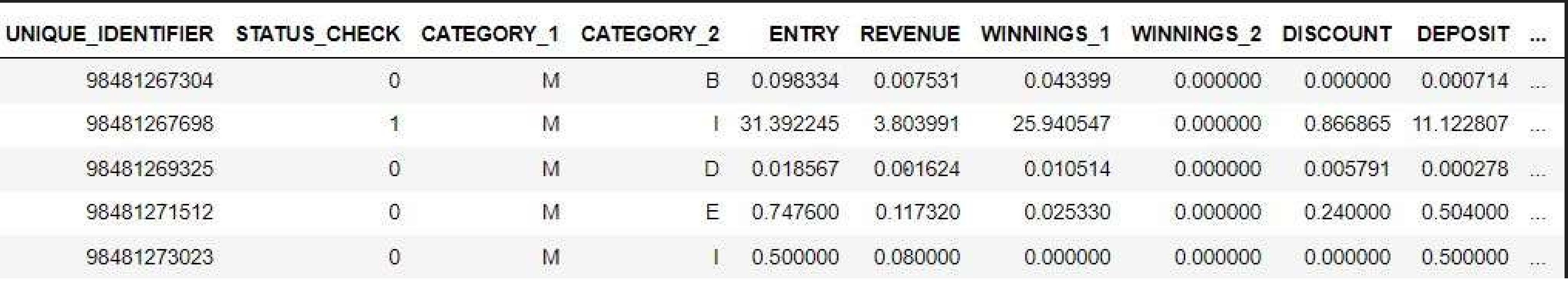


Data Preparation

Original



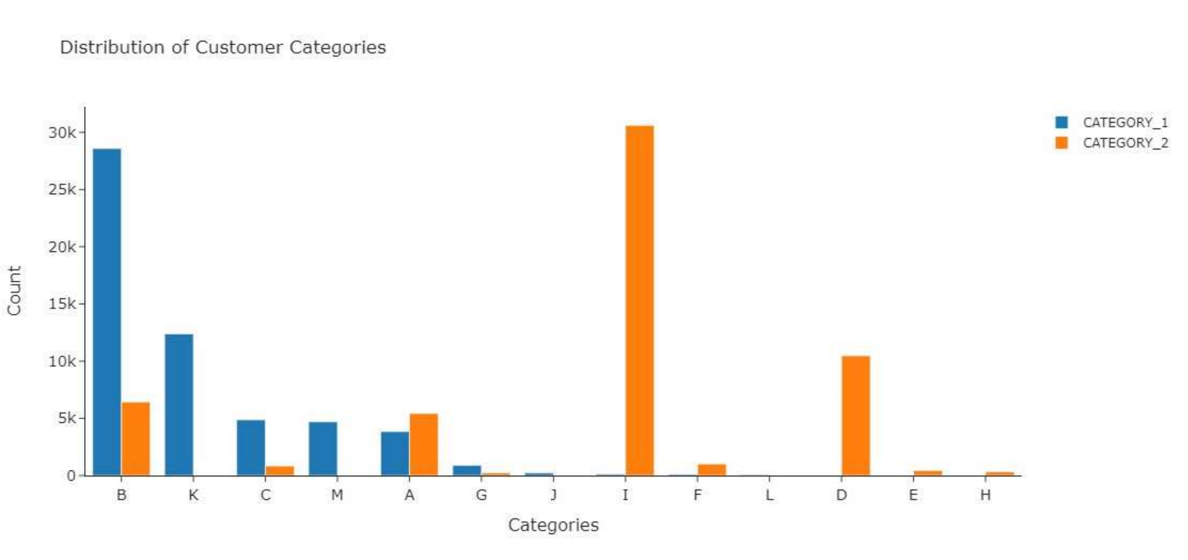
UNIQUE_IDENTIFIER	SEQUENCE_NO	STATUS_CHECK	CATEGORY_1	CATEGORY_2	ACTIVE_YN	ENTRY	REVENUE
98481267304	1	0	M	B	1	0.000000	0.000000
98481267304	2	0	M	B	1	0.137350	0.011550
98481267304	3	0	M	B	1	0.158350	0.010425
98481267304	4	0	M	B	1	0.444900	0.035850
98481267304	5	0	M	B	1	0.000000	0.000000
98481267304	6	0	M	B	1	0.000000	0.000000
98481267304	7	0	M	B	1	0.045050	0.002950



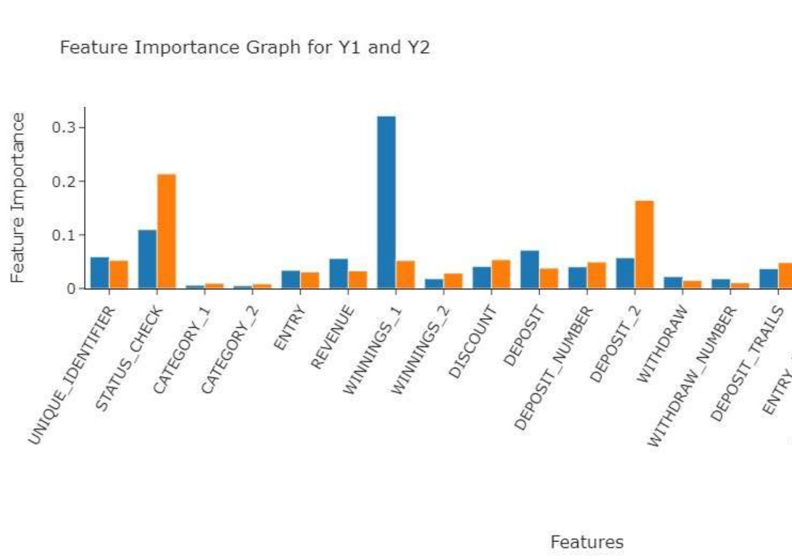
UNIQUE_IDENTIFIER	STATUS_CHECK	CATEGORY_1	CATEGORY_2	ENTRY	REVENUE	WINNINGS_1	WINNINGS_2	DISCOUNT	DEPOSIT	...
98481267304	0	M	B	0.098334	0.007531	0.043399	0.000000	0.000000	0.000714	...
98481267698	1	M	I	31.392245	3.803991	25.940547	0.000000	0.866865	11.122807	...
98481269325	0	M	D	0.018567	0.001624	0.010514	0.000000	0.005791	0.000278	...
98481271512	0	M	E	0.747600	0.117320	0.025330	0.000000	0.240000	0.504000	...
98481273023	0	M	I	0.500000	0.080000	0.000000	0.000000	0.000000	0.500000	...

Data Preprocessing

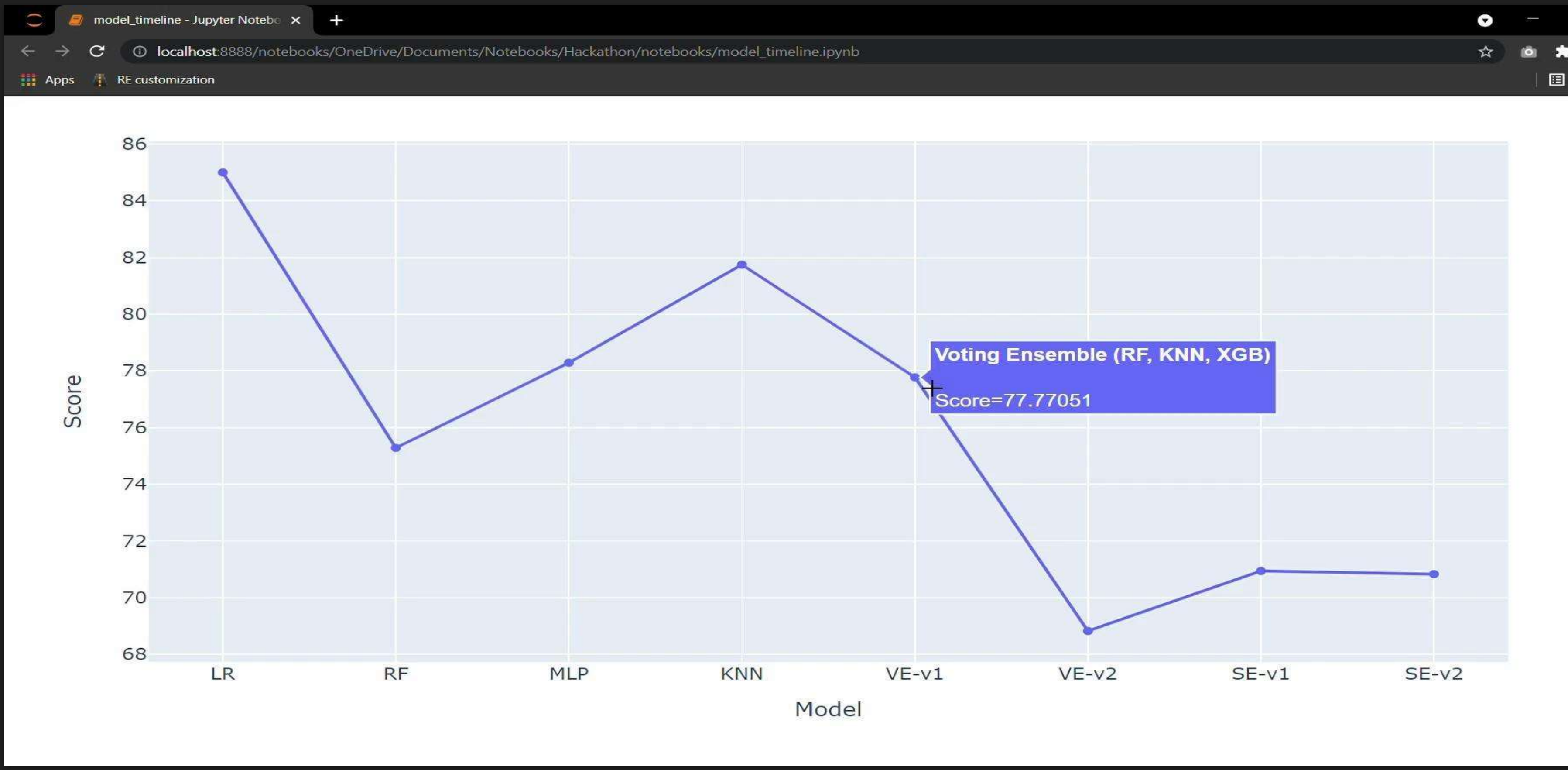
Categories Frequency Chart



Feature Importances (Tr



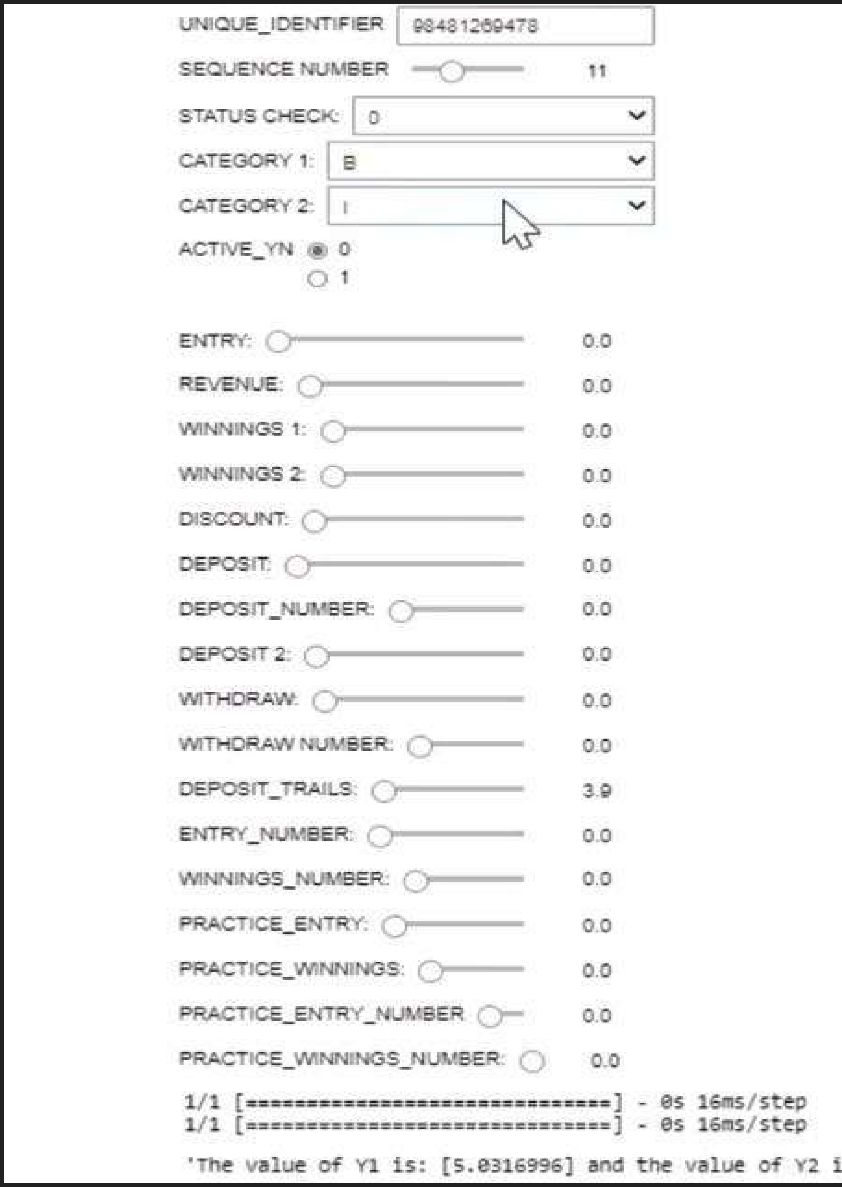
Methods Explored


model_timeline - Jupyter Notebo
localhost:8888/notebooks/OneDrive/Documents/Notebooks/Hackathon/notebooks/model_timeline.ipynb
Apps
RE customization
Voting Ensemble (RF, KNN, XGB)
Score=77.77051
Score
86
84
82
80
78
76
74
72
70
68
LR
RF
MLP
KNN
VE-v1
VE-v2
SE-v1
SE-v2
Model

Final Ensemble Model

Voting Regressor(
  Multilayer perceptron,
  Random forest,
  XGboost
)

- *An interactive visualization*



UNIQUE_IDENTIFIER 98481269478

SEQUENCE NUMBER 11

STATUS CHECK: 0

CATEGORY 1: B

CATEGORY 2: I

ACTIVE_YN ● 0 ○ 1

ENTRY: 0.0

REVENUE: 0.0

WINNINGS 1: 0.0

WINNINGS 2: 0.0

DISCOUNT: 0.0

DEPOSIT: 0.0

DEPOSIT_NUMBER: 0.0

DEPOSIT 2: 0.0

WITHDRAW: 0.0

WITHDRAW NUMBER: 0.0

DEPOSIT_TRAILS: 3.9

ENTRY_NUMBER: 0.0

WINNINGS_NUMBER: 0.0

PRACTICE_ENTRY: 0.0

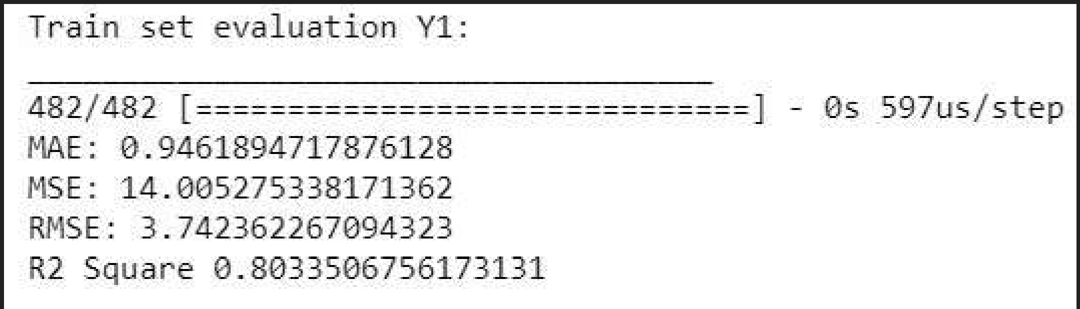
PRACTICE_WINNINGS: 0.0

PRACTICE_ENTRY_NUMBER 0.0

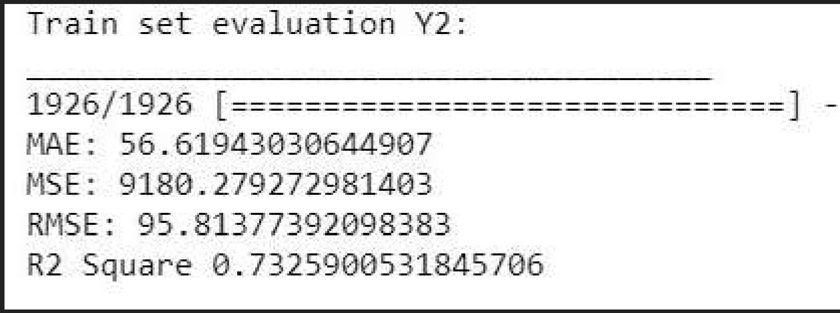
PRACTICE_WINNINGS_NUMBER: 0.0

```
1/1 [==============================] - 0s 16ms/step
1/1 [==============================] - 0s 16ms/step

'The value of Y1 is: [5.0316996] and the value of Y2 i
```

Results



```
Train set evaluation Y1:

____________________________________
482/482 [==============================] - 0s 597us/step
MAE: 0.9461894717876128
MSE: 14.005275338171362
RMSE: 3.742362267094323
R2 Square 0.8033506756173131
```



```
Train set evaluation Y2:

____________________________________
1926/1926 [==============================] -
MAE: 56.61943030644907
MSE: 9180.279272981403
RMSE: 95.81377392098383
R2 Square 0.7325900531845706
```

Epilogue

Conclusion

Various regression models such as RandomForest, Multilayer Perceptron and XGBoost v
trained on a processed dataset. Results showed that an ensemble of the best perform
models would yield the best score.

Challenges

Understanding the dataset in the context of its domain was a hurdle. Tackling the tem
component of the dataset was also a difficulty.

Future Scope

Using a deep learning approach for better feature selection and modelling